Listing of the Claims

1. (Currently Amended) A method of generating an index for a sequence that supports a non-contiguous subsequence match, comprising:

receiving a sequence;

receiving a window size;

encoding the sequence into a weighted-sequence;

encoding the weighted sequence into one or more one-dimensional sequences, wherein the length of each of the one or more one-dimensional sequences is less than the window size;

inserting each of the one or more one-dimensional sequences as one or more trie nodes into a trie structure; and

generating ~~the~~an index, wherein generating the index comprises~~comprising~~:

generating a current sequential ID and a maximum sequential ID pair for ~~generating~~each of the one or more trie nodes, wherein the current sequential ID of any descendant of a given trie node of the one or more trie nodes is between the current sequential ID of the given trie node and the maximum sequential ID;

generating an iso-depth link for each unique symbol in each of the one or more one-dimensional sequences, wherein the iso-depth link comprises trie nodes under the symbol; and

generating an offset list comprising an original position ~~of~~for each of ~~the~~ one or more subsequences in the weighted-sequence.

2

2. (Currently Amended) The method of claim 1, wherein encoding the sequence into ~~a~~the weighted-sequence comprises encoding the sequence with weights represented by real numbers[;].

3. (Currently Amended) The method of claim 2, wherein encoding the sequence with weights represented by real numbers[,] comprises discretizing the sequence into a number of equi-width units.

4. (Currently Amended) The method of claim 1, wherein inserting each of the one or more one-dimensional sequences into ~~a~~the trie structure ~~comprises~~is performed using a depth-first traversal.

5. (Currently Amended) The method of claim 1, further comprising creating ~~a~~the weighted-sequences index, ~~wherein the weighted-sequences index comprises an iso-depth index, comprises creating the weighted-sequences index,~~ wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a one-dimensional buffer.

6. (Currently Amended) The method of claim 1, further comprising ~~wherein~~ creating ~~a~~the weighted-sequences index, ~~wherein the weighted-sequences index comprises an iso-depth index, comprises creating the weighted-sequences index,~~ wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a B^+ tree.

7. (Currently Amended) The method of claim 1, ~~further comprising~~ wherein creating ~~the~~a weighted-sequences index, ~~wherein the weighted-sequences index comprises an iso-depth index, comprises creating the weighted-sequences index,~~ wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a linked list.

8. (Currently Amended) The method of claim 1, wherein receiving ~~a~~the sequence comprises receiving one or more elements in the sequence, wherein each of the one or more elements are represented by one or more ~~(symbol, weight)~~ pairs of symbol and weight elements.

9. (Currently Amended) The method of claim 8, ~~wherein receiving one or more elements in the sequence, wherein each of the one or more elements are represented by one or more (symbol, weight) pairs, and~~ wherein ~~each of~~ the symbol elements ~~of the one or more (symbol, weight) pairs~~ correspond to a non-uniform frequency distribution.

10. (Currently Amended) The method of claim 9, further comprising reordering the one or more one one-dimensional sequences using the non-uniform frequency distribution to generate a new sequence prior to inserting each of the one or more one-dimensional sequences into ~~a~~the trie structure ~~using the non-uniform frequency distribution to generate a new sequence.~~

4

11. (Currently Amended) The method of claim 10, wherein reordering the one or more one one-dimensional sequences using the non-uniform frequency distribution to generate a new sequence prior to inserting each of the one or more one-dimensional sequences into a~~the~~ trie structure ~~using the non-uniform frequency distribution to generate one or more new sequences~~, comprises:

(a) adding an offset 2*w*r to each of the weight element~~s of the one or more one-dimensional sequences~~ to generate a new weight, wherein w is a window size, and r is a frequency rank for a symbol of each of the symbol element~~sto generate a new weight~~;

(b) sorting the pairs of symbol and weight elements ~~the each element of the one or more one-dimensional sequences~~ by the new weight;

(c) placing a moving window of size 2*w*A on the ~~one or more~~ new sequences, wherein A is a frequency ~~the total number~~ of the symbols; and

(d) indexing the ~~one or more~~ new sequences in a new window.

12. (Currently Amended) The method of claim 1, wherein receiving a~~the~~ sequence comprises receiving one or more scientific datasets, transforming each of the one or more scientific datasets into one or more sequence~~s~~, and concatenating the one or more sequences to form a long sequence.

13. (Cancelled)

14. (Cancelled)

15. (Currently Amended) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform instructions executable by the machine~~machine-readable medium having instructions stored thereon for execution by a processor~~ to perform a method steps of generating an index for a sequence that supports a non-contiguous subsequence match, the method steps comprising~~the steps of~~:

receiving a sequence;

receiving a window size;

encoding the sequence into a weighted-sequence;

encoding the weighted sequence into one or more one-dimensional sequences, wherein the length of each of the one or more one-dimensional sequences is less than the window size;

inserting each of the one or more one-dimensional sequences as one or more trie nodes into a trie structure; and

generated an index, wherein generating the index comprises~~creating the index comprising~~:

generating a current sequential ID and a maximum sequential ID pair for each of the one or more trie nodes, wherein the current sequential ID of any descendant of a given trie node of the one or more trie nodes is between the current sequential ID of the given trie node and the maximum sequential ID;

generating an iso-depth link for each unique symbol in each of the one or more one-dimensional sequences, wherein the iso-depth link comprises trie nodes under the symbol; and

generating an offset list comprising an original position ~~of~~ <u>for</u> each of ~~the~~ one or more subsequences in the weighted-sequence.

16. (Cancelled)